

# Unidata and Data Portals: Community, THREDDS, netCDF

Ethan Davis  
UCAR/Unidata



# UCAR/Unidata

- Community: University geoscience researchers and educators (large and small institutions)
- Distribution of real-time meteorological data (IDD/LDM)
- Community is used to having analysis and display power on desktop (thick clients)
- A number of Unidata sites host web portals (real-time data, static set of plots)

# Desired Features/Capabilities

- Usable by Unidata community members
  - Can easily be run by small groups
    - Minimal system admin requirements
    - Minimal equipment requirements.
  - Capabilities and flexibility for power users
- Support for generation of metadata (i.e., doesn't require much knowledge of metadata standards).
- Flexible in terms of data/metadata backend storage (file, database, etc.)

# Desired Features/Capabilities

- Search and Discovery Capabilities
  - Browse
  - Keyword on metadata content
  - Spatial
  - Temporal
  - Parameters

# Desired Features/Capabilities

- Not just HTML based interface. Stand-alone applications need to interact with data portal services
- Need programmatic interfaces (e.g., web services) to data portal service:
  - Search and discovery service
  - Visualization service
  - Access service

# THREDDS Update: Catalogs

- THREDDS catalogs for communicating information about a collection of datasets
- Simple structure:
  - Datasets can contain datasets (hierarchical structure)
  - Datasets have zero or more access methods
    - Several encodings to simplify common cases
  - Datasets can contain (or reference) metadata
- Referenced by publications and educational materials

# THREDDS Update: Catalogs

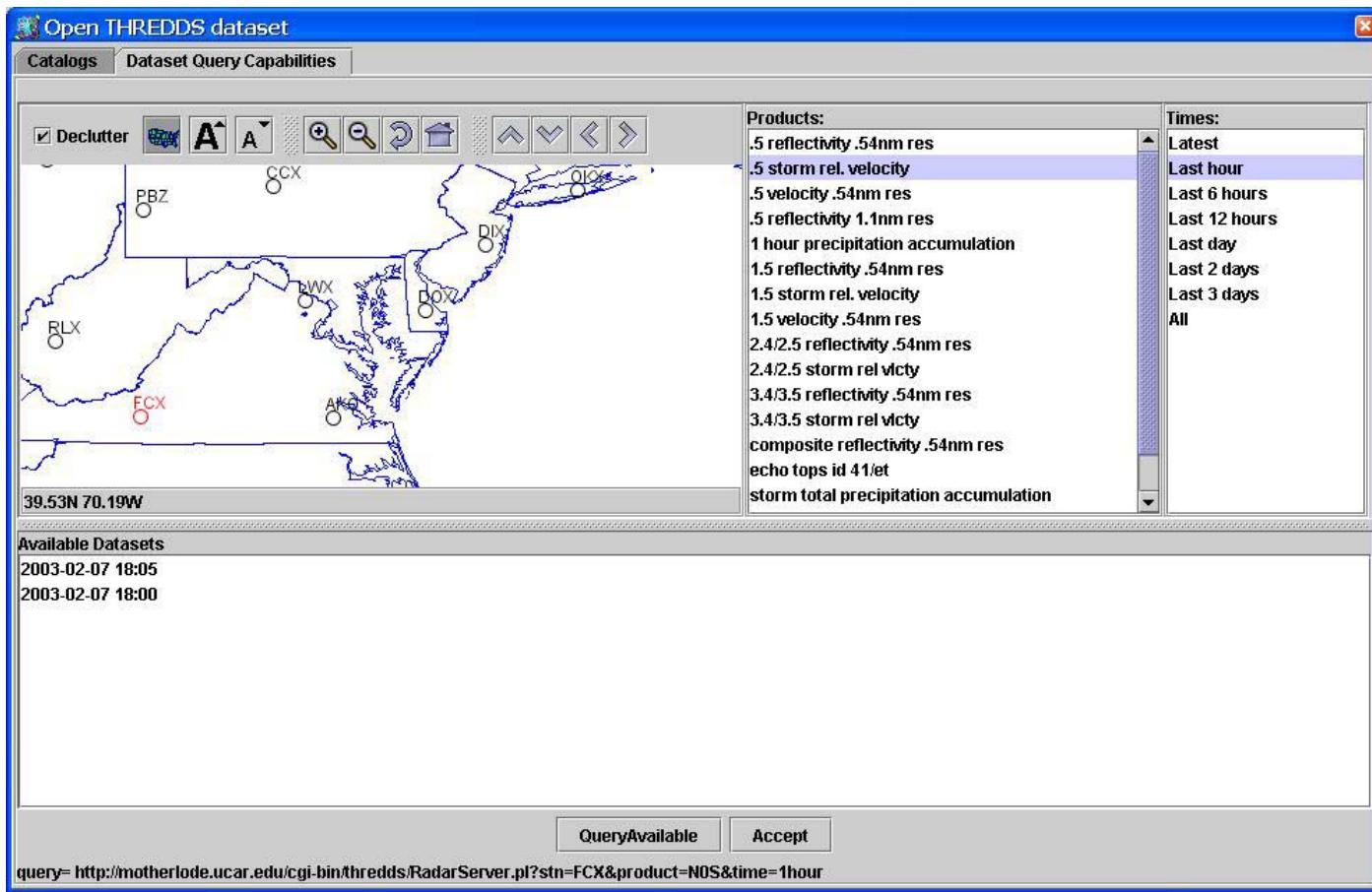
```
<catalog version="0.6">
  <dataset name="Unidata IDD Model Data">
    <service name="mlode" serviceType="DODS"
      base="http://motherlode/cgi-bin/nph-dods/" />
  <dataset name="NCEP Eta 80km CONUS model data ">
    <documentation>
      NCEP Eta model runs from the last five days on an 80km CONUS grid (211).
    </documentation>
    <metadata metadataType="DublinCore"
      xlink:href="http://motherlode/dc/eta_211.dc.xml" />
    <dataset name=" NCEP Eta 80km CONUS 2003-09-24 00Z"
      serviceName="mlode" urlPath="2003092400_eta_211.nc"/>
    <dataset name=" NCEP Eta 80km CONUS 2003-09-24 12Z"
      serviceName="mlode" urlPath="2003092412_eta_211.nc"/>
  ...
...
```

# THREDDS Update: Discovery

- Discovery services (browse, search/subset)
  - Digital libraries harvesting metadata
    - NSDL, DLESE via OAI (Open Archive Initiative) Protocol
    - GCMD
    - Consider granularity
  - Digital library search
    - OAI search
    - Others (e.g., GCMD search)
  - THREDDS DQC (Dataset Query Capabilities)
    - Specify set of queries allowed for a dataset collection
    - Service responds with a catalog

# THREDDS Update: Discovery

THREDDS DQC: subsetting/searching large or rapidly changing collections



# THREDDS Update: Metadata

- Metadata
  - Dublin Core, ADN, DIF, etc
  - Spatial and temporal ranges (limits, density, etc)
  - Parameters (CF to GRIB to GCMD, etc)
  - Tools for harvest and creation of metadata
  - Looking at GIS standards
  - NcML
    - Coordinate Systems
    - Dataset: subset, modify, aggregate
    - GIS

# THREDDS Update: Access

- THREDDS WCS (Web Coverage Service)
  - Gridded datasets
  - GeoTIFF
  - Data model mismatch
    - GIS vs Scientific Data
    - 2D vs 3- and 4-D
    - Time dimension(s)
    - Weird spatial dimensions, e.g., trajectories, x-z plots
    - Can't serve any dataset with WCS

# THREDDS Update: Future

- THREDDS Priorities
  - Update catalog schema
  - Update DQC schema
  - THREDDS WCS
  - Tools for creation/harvest of metadata
  - Digital libraries harvesting metadata
  - ...
- THREDDS and ESP
  - What are ESP priorities?
  - What can THREDDS do for/with ESP? When needed?

# NetCDF

- Looking into using Java netCDF API on top of GRIB/GRIB2
- “Merging” NetCDF and HDF: Unidata has been funded to work with NCSA to create the next generation netCDF API using HDF5 as the storage layer